

© 2012 Sai Zhang

SYSTEM-LEVEL OPTIMIZATION OF THE DC-DC VOLTAGE  
REGULATOR AND CORE FOR SUB/NEAR-  
THRESHOLD VOLTAGE OPERATION

BY

SAI ZHANG

THESIS

Submitted in partial fulfillment of the requirements  
for the degree of Master of Science in Electrical and Computer Engineering  
in the Graduate College of the  
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Adviser:

Professor Naresh R. Shanbhag

# ABSTRACT

Switched capacitor voltage regulator module (SC-VRM) is suitable for low power embedded systems operating in near/sub-threshold region due to its high conversion ratio and compactness. However, existing optimization for SC-VRM is separated from the embedded core design and therefore leads to sub-optimal system energy efficiency. In this thesis, we propose to jointly optimize the switched capacitor voltage regulator module (SC-VRM) and the compute core to minimize system energy per instruction. A core-aware SC-VRM energy model is developed and employed to solve the joint optimization problem. We also propose and optimize a reconfigurable SC-VRM architecture. Simulation results in a 130nm CMOS process indicate that the core-aware SC-VRM model predicts energy from circuit simulations to within 5%, and that the proposed approach results in a maximum system energy savings of 8% to 38.9%. The reconfigurable SC-VRM achieves 15% to 52% energy saving as compared to an efficiency-optimized design.

*To my parents and my friends for their love and support*

# ACKNOWLEDGMENTS

I would like to thank my advisor, Professor Naresh R. Shanbhag, for his continuous guidance and invaluable comments. I also would like to thank my family for their love and all my friends for their support.

# TABLE OF CONTENTS

CHAPTER 1	INTRODUCTION . . . . .	1
1.1	Motivation . . . . .	1
1.2	Background . . . . .	3
1.3	Thesis Organization . . . . .	12
CHAPTER 2	SC-VRM PRINCIPLE OF OPERATION AND EN- ERGY MODEL . . . . .	15
2.1	SC-VRM Principle of Operation . . . . .	15
2.2	SC-VRM Energy Model . . . . .	16
CHAPTER 3	JOINT OPTIMIZATION OF THE SC-VRM AND THE CORE . . . . .	24
3.1	Core-aware SC-VRM Energy Model . . . . .	25
3.2	System-level Energy Optimization . . . . .	27
3.3	Simulation Results . . . . .	31
CHAPTER 4	CONCLUSION . . . . .	37
REFERENCES	. . . . .	38

# CHAPTER 1

## INTRODUCTION

### 1.1 Motivation

Moore's law has driven the continuous scaling of transistors for the past several decades [1]. We have entered the age of cloud computing [2], where computing is becoming increasingly intelligent and ubiquitous. As shown in Figure 1.1 [2], cloud computing comprises several layers of information processing capability. In the core, there are data centers and servers which provide high throughput for the user. The second layer consists of mobile devices that not only perform local computation, but also act as gateways to the high performance data centers. Finally, in the outermost layer, a tremendous number of sensor nodes acquire information from the environment and pass the data to the core. The design of the portable sensor nodes is critical for the pervasiveness of computing, and also challenging due to the strict form factor and power consumption requirements.

Energy per instruction (EPI) has to be minimized through careful design to extend system lifetime for sensor nodes with limited power supply. In the past, voltage scaling has been the dominant technique for energy reduction. However, the supply voltage has stopped scaling since 65 nm technology node as shown in Figure 1.2. The increased leakage in advanced technologies has made power a primary concern in today's integrated circuit designs [3]. As shown in [4], [5], due to the tradeoff between active and leakage energies, the minimum energy operating point (*MEOP*) is usually achieved in the near or sub-threshold region where the supply voltage is several hundreds of millivolts. Nevertheless, the battery used to supply power to the portable nodes usually has a voltage ranging from 1.2 V to 3.6 V [6]. In sensor nodes where small form factor is necessary, a thin film battery with an output voltage of around 4 V can be used [7]. This gap between battery output voltage and



Figure 1.1: The concept of cloud computing [2].

the desired supply voltage requires a high conversion ratio voltage regulator module (VRM). The current approach is to address the design of the VRM and the compute core separately, where the VRM is designed to provide a nearly constant supply voltage to the core, and the core operates at mostly fixed clock frequency. The requirement of a high conversion ratio and tight regulation under low output voltage puts a severe burden on VRMs. Among the three types of commonly used VRM topologies, the SC-VRM can achieve a high conversion ratio and is relatively easy to integrate [8]. However, conventional SC-VRM suffers from charge transfer loss, drive loss, and bottom plate capacitor loss. Design optimization is thus needed to design the VRM with the compute core in order to achieve system energy-efficiency.

In this thesis, we propose a core-aware energy model of the SC-VRM, based on which the joint optimization problem is formulated including the design parameters of both the SC-VRM and the core. This optimization problem is solved in a structured manner to demonstrate the benefits of the proposed joint optimization. Finally, we propose and optimize a reconfigurable SC-VRM architecture.



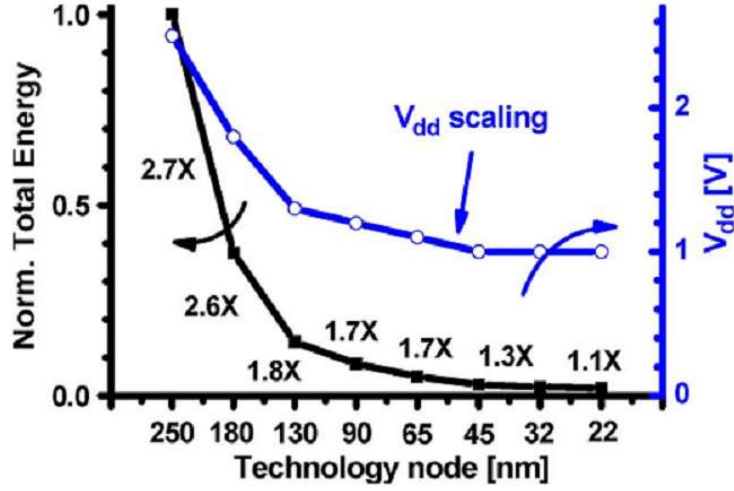


Figure 1.2: Technology scaling trends for supply voltage and energy[3].

## 1.2 Background

### 1.2.1 Energy-optimal computing

It has been shown that CMOS circuits can function correctly with a supply voltage far below the threshold voltage  $V_{th}$ . However, as pointed out in [4], due to the tradeoff between dynamic energy and leakage energy, a *MEOP* exists in near/sub-threshold region. We will use an inverter chain as an illustrative example.

Assuming an  $n$  stage inverter chain, the activity factor, load capacitance per stage, supply voltage, leakage current, and clock period are denoted as  $\alpha$ ,  $C_{core}$ ,  $V_{dd}$ ,  $I_{leak}$ , and  $T_{clk}$ , respectively. We assume that clock frequency is always the maximum achievable frequency, which is equal to the inverse of the critical path delay. The dynamic energy, leakage energy and total energy per clock cycle can be expressed as:

$$E_{dyn} = n\alpha C_{core} V_{dd}^2 \quad (1.1)$$

$$E_{leak} = I_{leak} V_{dd} T_{clk} = nC_{core} V_{dd}^2 (n\eta e^{-\frac{V_{dd}}{mV_T}}) \quad (1.2)$$

$$E_{core} = E_{dyn} + E_{leak} = nC_{core} V_{dd}^2 (\alpha + n\eta e^{-\frac{V_{dd}}{mV_T}}) \quad (1.3)$$

Figure 1.3 shows the dynamic, leakage, and total energy as a function of  $V_{dd}$  in a 50 stage inverter chain using TSMC 180 nm process with  $\alpha = 0.2$ . As shown in Figure 1.3, dynamic energy reduces with supply voltage due to the quadratic relationship between dynamic energy and voltage, but leakage energy increases with lower supply voltage in sub-threshold region, primarily due to the exponential increase in critical path delay. Therefore, an *MEOP* exists which balances the dynamic energy and the leakage energy. We define *MEOP* via a 3-tuple  $(E^*, V_{dd}^*, T^*)$ , where  $E^*$ ,  $V_{dd}^*$  and  $T^*$  are the energy consumption, supply voltage and clock period, respectively at the *MEOP*. It is clearly seen from Figure 1.3 that *MEOP* exists in the sub-threshold region. The *MEOP* is a function not only of technology parameters, but also of logic depth and activity factor. Higher logic depth will increase  $V_{dd}^*$  because more transistors contribute to leakage energy, while higher activity factor will decrease  $V_{dd}^*$  because dynamic energy is higher.

#### 1.2.1.1 Practical *MEOP* in different applications

The analysis in the previous section assumes there are no task deadlines. This enables a circuit to run arbitrarily slowly. Real applications can be divided into two categories, throughput unconstrained applications (TUA) and throughput constrained applications (TCA). TUAs do not have a deadline; typical examples are medical or sensor applications. TCAs have performance requirements and have to meet delay specifications. Such applications can range from processors to mobile CODECs. *MEOPs* exist in both cases but they are not identical, as shown in Figure 1.4. For TUAs, the circuit can always run at the voltage that has the lowest energy per instruction, regardless of the speed. The  $MEOP_{ntc}$  is thus the same as previously derived in

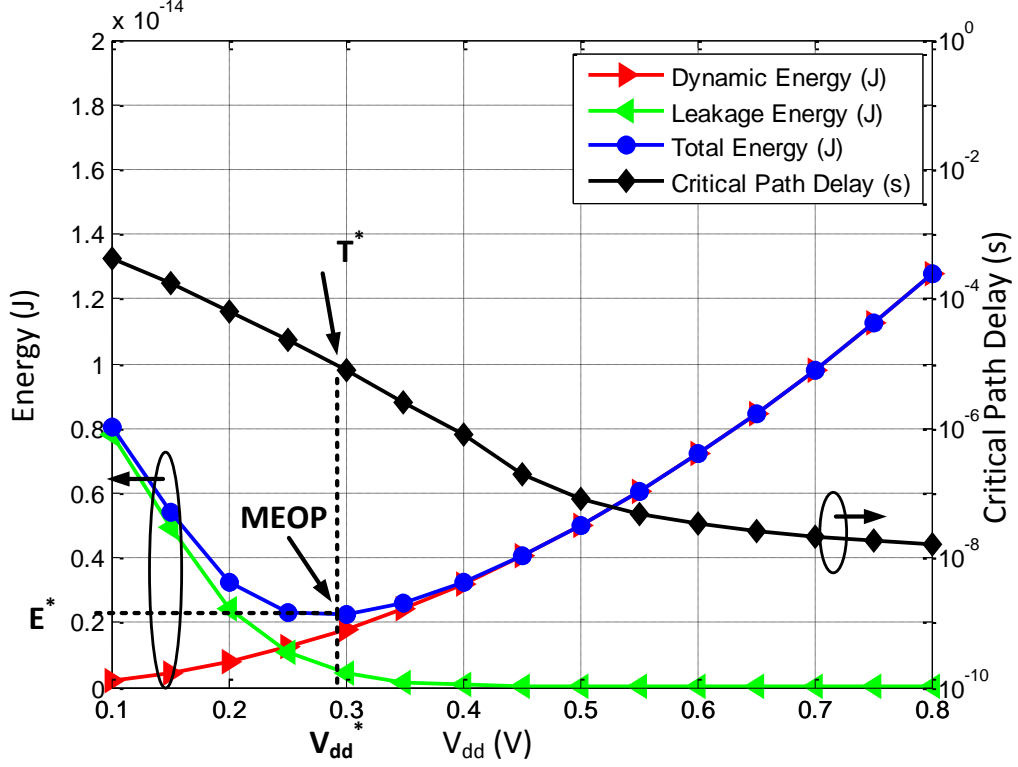


Figure 1.3: MEOP of a 50 stage inverter chain in a 180 nm CMOS process.

Figure 1.3. For TCAs, it is usually not possible to reduce the voltage down to the *MEOP* voltage since the delay requirement will be violated. In this case, the *MEOP* voltage is the lowest  $V_{dd}$  under which critical path delay meets the performance requirements.

#### 1.2.1.2 The *MEOP* for TUAs

For a fixed circuit architecture and activity factor, the *MEOP* is usually fixed. However, circuit design techniques such as parallelization and pipelining will impact the *MEOP*. Circuit designers can use architectural optimization to further reduce energy consumption at the *MEOP*.

For TUAs, parallelization does not affect  $MEOP_{TUA}$  because parallelization uses identical circuit blocks to enhance throughput. Each block has the same architecture with respect to the original one, so the  $MEOP_{TUA}$  will not be affected by parallelization. On the other hand, the pipelining level  $N$  does modify the  $MEOP_{TUA}$ . Figure 1.5 shows the effect of the pipelining level  $N$  on the  $MEOP_{TUA}$ . The same 50 stage inverter chain as in Figure 1.3 is used

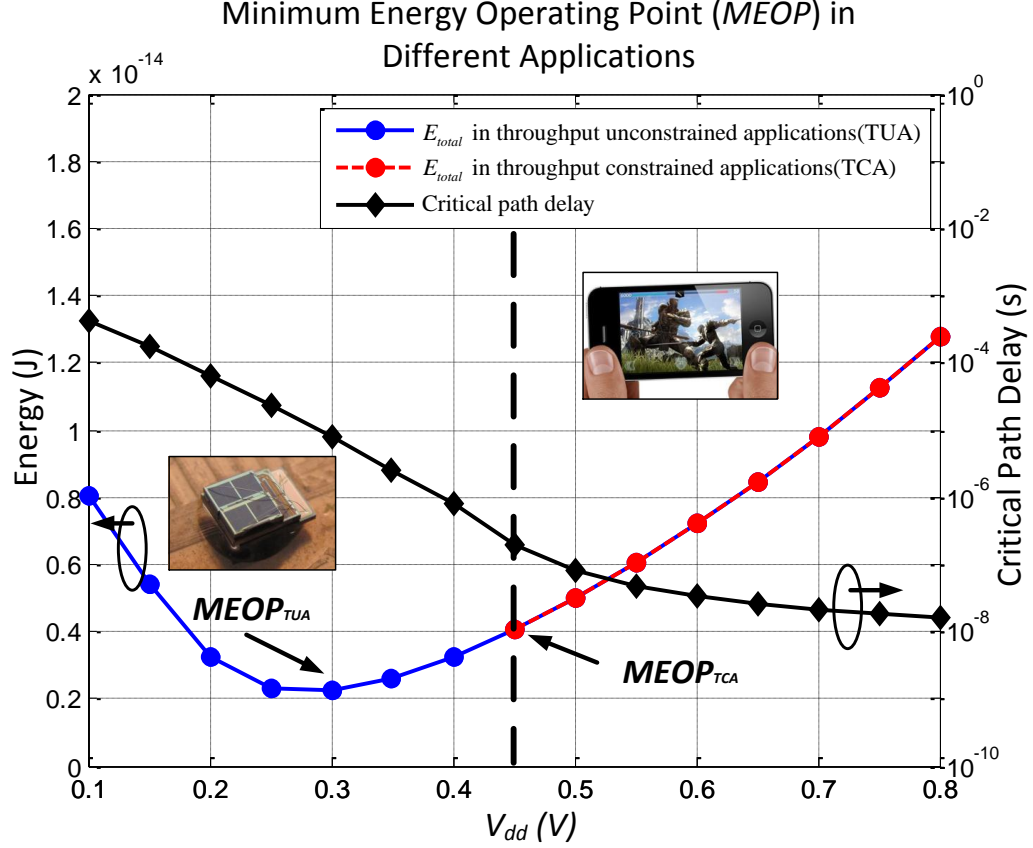


Figure 1.4: *MEOP* for TUAs and TCAs, a simple 50 stage inverter chain in 180nm process is simulated to obtain its *MEOP*.

in simulation and pipeline stages of  $N = 2, 5$  and  $10$  are assumed in order to evaluate their effect on the  $MEOP_{TUA}$ . With more pipelining stages, critical path delay is reduced, which makes dynamic energy increasingly dominant.  $V_{dd}^*$  therefore moves to lower voltages as pipelining level increases. In practice, pipelining will introduce energy overhead, which implies the existence of an optimum pipelining stage. It is reported in [9] that optimum pipelining occurs in sub-threshold region with a register/logic area ratio of  $0.2-0.3$ . This *MEOP* is the global minimum energy point to achieve error-free operation.

#### 1.2.1.3 The *MEOP* for TCAs

Figure 1.6 shows the effect of parallelization and pipelining on  $MEOP_{TCA}$ . To obtain this figure, the same 50 stage inverter chain as in Figure 1.3 is employed. Parallelization factor ( $P$ ) and pipelining levels ( $N$ ) are swept

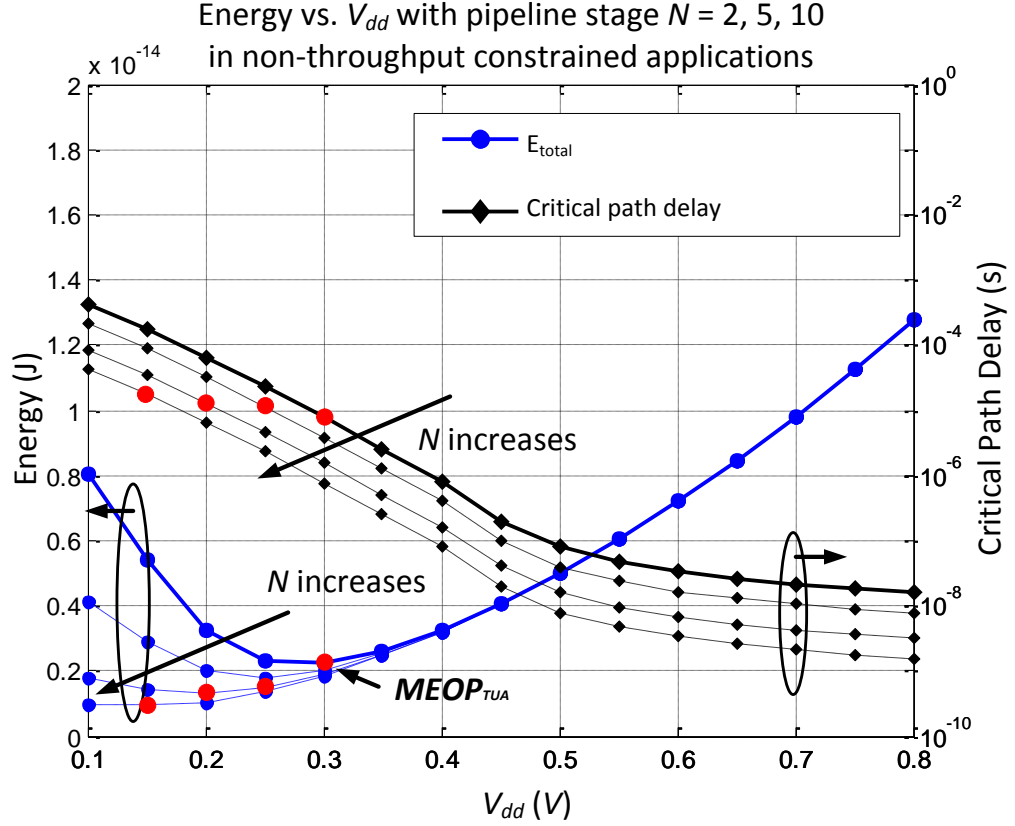


Figure 1.5: Effect of the pipelining level  $N$  on  $MEOP_{TUA}$ . Pipelining reduces the  $MEOP$  voltage  $V_{dd}^*$  and the corresponding energy per instruction  $E^*$ . The pipelining levels evaluated are  $N = 1, 2, 5$  and  $10$ .

to evaluate their effects on the  $MEOP_{TCA}$ . As shown in Figure 1.6(a), for TCAs, parallelization will affect the  $MEOP_{TCA}$ , although the energy vs.  $V_{dd}$  relationship does not change with parallelization factor. The reason is that for TCAs, higher parallelization factor  $P$  will reduce the delay requirement.  $P$ -level parallelization will allow each circuit block to run  $P$  times slower while delivering the same performance. The relaxed delay requirement can be used to reduce voltage for energy reduction. Note that when the parallelization factor is large, each circuit block can operate at the  $MEOP_{TUA}$  and meet the performance requirement. In this case  $MEOP_{TUA}$  and  $MEOP_{TCA}$  are the same. Pipelining will reduce the  $V_{dd}^*$  to lower voltages as well as relax the delay requirement. Therefore, the  $MEOP_{TCA}$  will reduce to lower voltages as the pipelining stage increases, as shown in Figure 1.6(b).

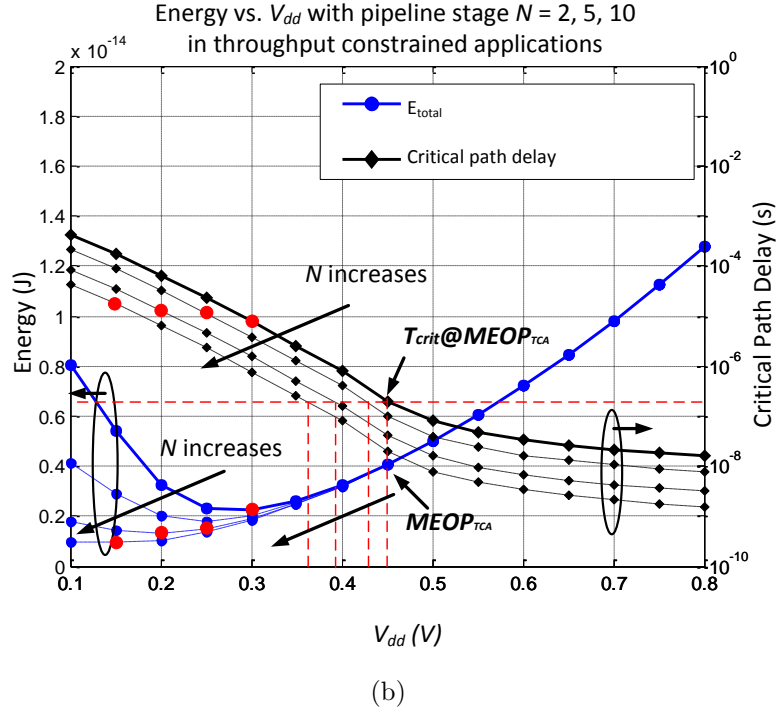
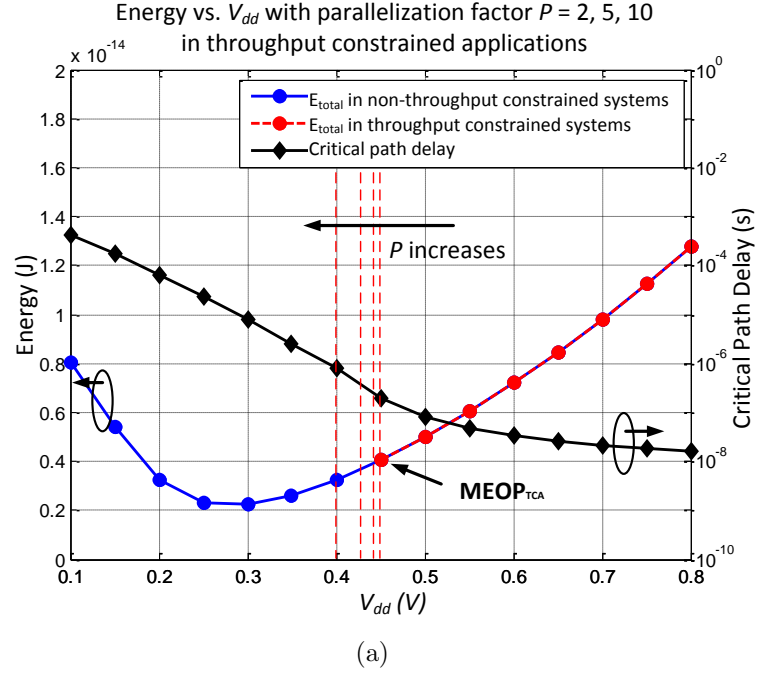


Figure 1.6: Effect of parallelization and pipelining on  $MEOP_{tc}$ : (a) parallelization relaxes the delay requirement, enabling the reduction of  $MEOP$  voltage and corresponding energy per instruction; (b) pipelining also reduces the  $MEOP$  voltage and the corresponding energy per instruction. The parallelization factors ( $P$ ) and pipeline stages ( $N$ ) evaluated are 1, 2, 5 and 10.

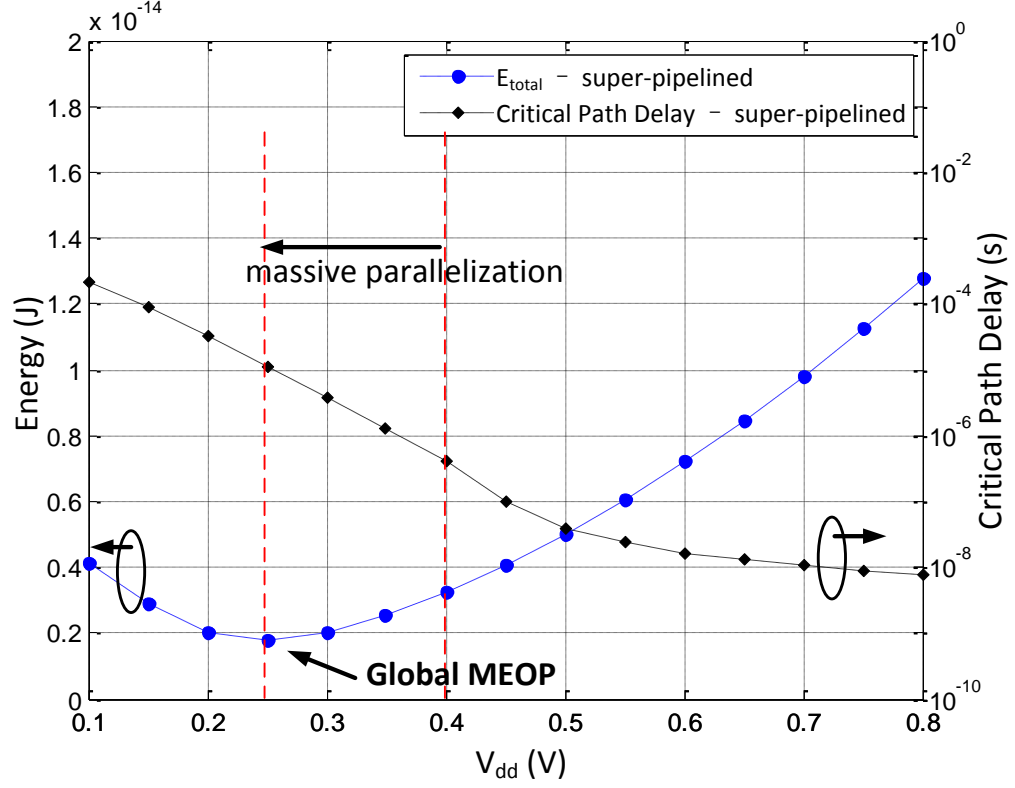


Figure 1.7: Global *MEOP* can be achieved via super-pipelining to reduce  $MEOP_{ntc}$  combined with massive parallelization to enhance throughput.

#### 1.2.1.4 Energy-optimal computing

The facts that pipelining can move the  $MEOP_{TUA}$  towards the global minimum energy point, and that parallelization can enable the circuit to operate at the  $MEOP_{TUA}$  while meeting the performance requirement, imply that the combination of these two techniques will achieve energy-optimality for a given throughput requirement. As shown in Figure 1.7, energy-optimal computing is achieved by first using super-pipelining [9] to shift the *MEOP* to a global *MEOP*, and then using massive parallelization to enhance throughput. This method, while appealing in terms of energy efficiency, is seldom used in practice, because the required value of  $P$  to deliver higher performance at the global *MEOP* is large, which will violate area and cost constraints in real applications.

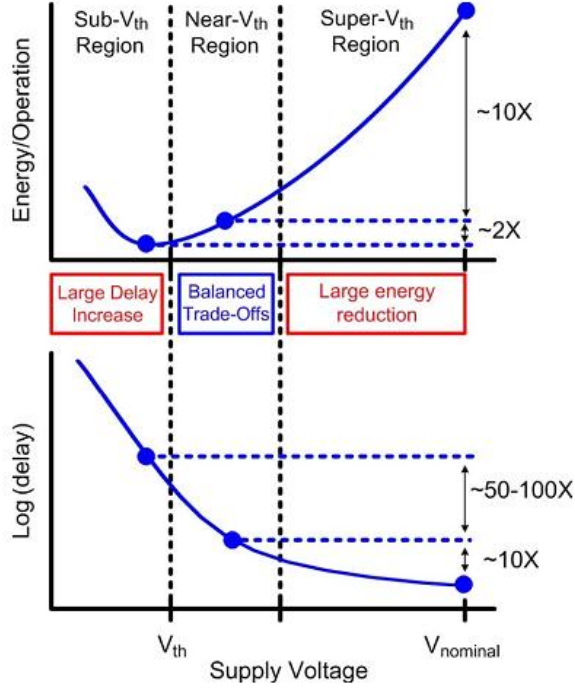


Figure 1.8: Energy and delay in different supply voltage operating regions [3].

#### 1.2.1.5 Existing challenges and recent research directions

Operating the circuit in sub-threshold *MEOP* is an attractive solution for energy-efficient computing. However, several challenges still exist in this area, as illustrated in [3]. Among the barriers are severe performance penalty due to increased delay, increased process variation, and orders of magnitude higher device failure. To address these problems, [3] suggests operating at near-threshold instead of deep sub-threshold due to lower performance degradation and better robustness. The motivation for near-threshold computing can be derived Figure 1.8 [3]. As the operating point of a design moves from super-threshold to near-threshold, we can get 10X energy benefits with relatively small performance penalty. Further reduction of supply voltage yields only marginal energy reduction of around 2X while the performance degradation is 50-100X due to the increased delay and above mentioned issue associated with sub-threshold computing. So an operating point which balances performance requirement and energy reduction should exist in the near-threshold region.



### 1.2.2 DC-DC converter basics

There are several important design metrics for DC-DC converters. Line regulation is defined as the change in output voltage given a small change in input voltage:

$$R_{line} = \frac{\partial V_{out}}{\partial V_{in}} \quad (1.4)$$

Load regulation, also known as output impedance, is defined as the change in output voltage given a small change in load current:

$$R_{load} = \frac{\partial V_{out}}{\partial I_{out}} \quad (1.5)$$

Efficiency is another important design metric for DC-DC converter, and is defined as the ratio of output power to input power:

$$\eta = \frac{P_{out}}{P_{in}} \quad (1.6)$$

There are three commonly used DC-DC converter topologies: 1) linear regulator, 2) switching converter, and 3) switched capacitor voltage regulator module (SC-VRM), as shown in Figure 1.9. The linear regulator (Figure 1.9(a)) uses high gain amplifier and series-shunt feedback to regulate the voltage to a desired reference level. The line regulation for a linear regulator is determined by the voltage reference characteristic, which is usually very good with supply independent biasing. With sufficiently high loop gain, the load regulation and dynamic response time can be very small. A major problem with using a linear regulator in sub/near-threshold region is that its efficiency is directly determined by the ratio of  $V_{out}$  to  $V_{in}$ , which is low for near/sub-threshold circuits.

A switching converter (Figure 1.9(b)) is most commonly employed in high performance applications due to its high efficiency and power density. This circuit uses duty cycle controlled switches to convert a DC voltage into a pulse train, followed by an LC low-pass filter to extract its DC component. Since an ideal LC filter is lossless, the maximum achievable efficiency is 100% regardless of conversion ratio. Given a high loop gain, a switching converter can have a good line and load regulation. With proportional-integral (PI) control, the reference tracking error can be eliminated. Advanced control schemes such as current control can also be used for better dynamic response

and stability. The problem with using a switching converter in sub/near-threshold region is the form factor issue because the inductor is difficult to integrate on chip. Another issue associated with using a switching converter under high conversion ratio is that the control to output transfer function becomes more sensitive to control variations as the duty cycle reduces, so the output voltage becomes more prone to control errors.

Finally, SC-VRM employs a capacitor array to store and transfer charge. Voltage conversion is achieved via topology change. Since the open loop output voltage is directly related to both input voltage and load current, sufficient loop gain is necessary to achieve good line and load regulation. The advantages of using SC-VRM in sub/near-threshold applications are its compactness and a high conversion ratio. As pointed out in [10], SC-VRM demonstrates a single pole transfer function and is stable without the need for complex compensation as required in the previous two cases. In terms of efficiency, SC-VRM can have a high efficiency when the output voltage is close to the ideal conversion output. Designers can choose appropriate converter topology and optimize it to achieve high efficiency under high conversion ratio. Therefore, the SC-VRM is most suitable to sub/near-threshold applications among the three topologies in Figure 1.9.

### 1.3 Thesis Organization

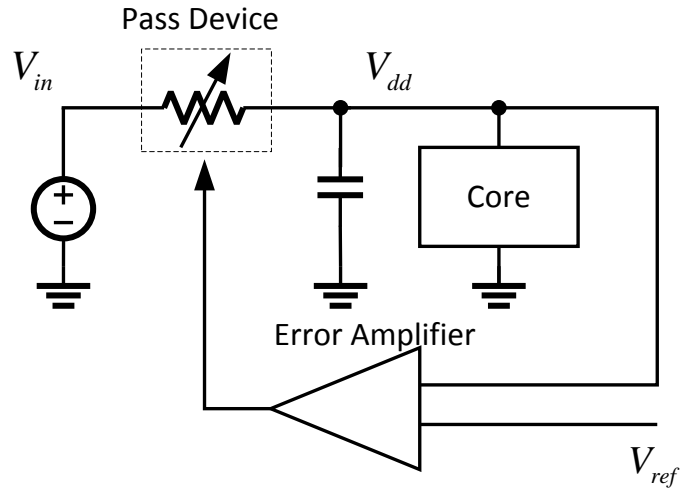
The organization of the thesis is as follows. Chapter 1 presented the importance and challenges for designing low power, low to medium throughput sensor nodes. Background for minimum energy operating point design as well as basic DC-DC converter architecture was presented. It was illustrated that SC-VRM, due to its high conversion ratio and compactness, is a feasible architecture for low power applications.

Chapter 2 focuses on design and optimization of SC-VRM, specifically for low power applications. The principle and an analysis method for SC-VRM are presented. An energy model for SC-VRM is developed, based on which optimization for different load conditions is carried out.

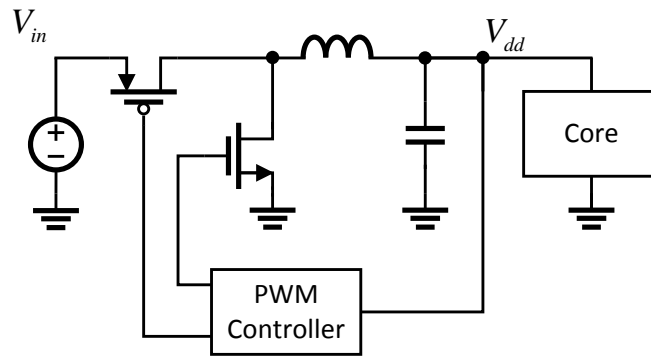
Chapter 3 presents the core-aware SC-VRM model, formulates and solves the joint optimization problem. Simulation results are presented to show potential benefits of system optimization. In addition, we extend the system

optimization to reconfigurable SC-VRM architectures.

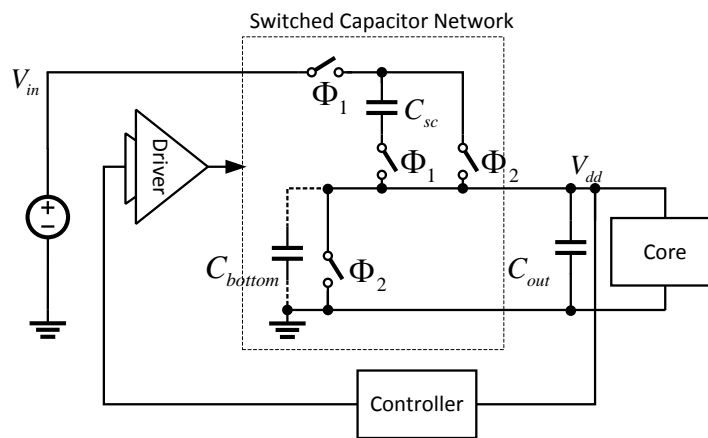
Chapter 4 concludes and addresses future directions.



(a)



(b)



(c)

Figure 1.9: Three commonly used DC-DC converter topologies: (a) linear regulator, (b) switching converter, and (c) SC-VRM.

## CHAPTER 2

# SC-VRM PRINCIPLE OF OPERATION AND ENERGY MODEL

### 2.1 SC-VRM Principle of Operation

The SC-VRM utilizes a capacitor array to store and transfer charge from the input to the output. Voltage conversion is achieved by modifying the topology of the capacitor array. The fundamental reason for the higher efficiency of SC-VRM compared to a linear regulator is because the former employs charge recycling. In a linear regulator, the input and the output are always connected, so the charge drawn from the input can be utilized to perform computation only once. In SC-VRM, since charge is stored inside the capacitor array in one phase and reused in subsequent phases, the same amount of charge can perform multiple computations. As a result, less energy is consumed by the SC-VRM to deliver the same performance as a linear regulator.

This concept of charge recycling not only helps to understand the operation of SC-VRM, but also serves as a convenient tool to analyze a SC-VRM. There are many different SC-VRM topologies, and [11] gives a good summary. These topologies, though quite different, can be analyzed using the charge multiplier method [11] which is based on the concept of charge recycling. A charge multiplier denotes the magnitude and sign of the charge flowing in or out of a capacitor during each phase. By analyzing the configuration of a SC-VRM in each phase and solving for the corresponding charge multipliers, the conversion ratio of the SC-VRM can be obtained.

A Fibonacci SC-VRM is as used an example to illustrate this method. A Fibonacci SC-VRM shown in Figure 2.1(a) can achieve the highest conversion ratio given the same number of charge transfer capacitors. Figure 2.1(b) and 2.1(c) show the configuration in each of the two phases. The charge multiplier for each capacitor can be determined by applying charge

Table 2.1: Charge multiplier for Fibonacci SC-VRM

	$\beta_{in}$	$\beta_{C1}$	$\beta_{C2}$	$\beta_{out}$
Charge multiplier in $\phi_1$	$q$	$q$	$q$	$2q$
Charge multiplier in $\phi_2$	$0$	$-q$	$-q$	$q$

conservation to each capacitor over one period, i.e., after the completion of two phases. Table 2.1 shows the charge multiplier for each capacitor. The ideal conversion ratio can be found by equating the total input and output energy:

$$\frac{V_{out}}{V_{in}} = \frac{\beta_{in}}{\beta_{out}} = \frac{1}{3} \quad (2.1)$$

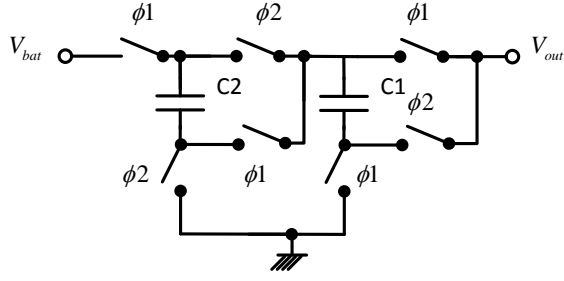
Here the conversion ratio is the ideal ratio when no energy loss is assumed. In practice, the output voltage will always be lower than the ideal case due to the intrinsic losses associated with the SC-VRM.

## 2.2 SC-VRM Energy Model

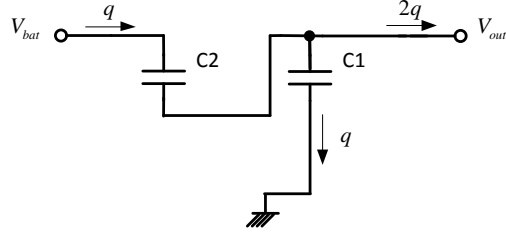
SC-VRM transfers charge by altering the configuration of the capacitor array. This mechanism causes energy loss due to the charge sharing between the capacitors with different voltages. Moreover, to change the configuration of the capacitors, driver circuits are needed to switch on/off the power switches during phase change, which contributes to additional losses. Other loss components also need to be identified before systematically optimizing the SC-VRM.

### 2.2.1 Charge transfer loss

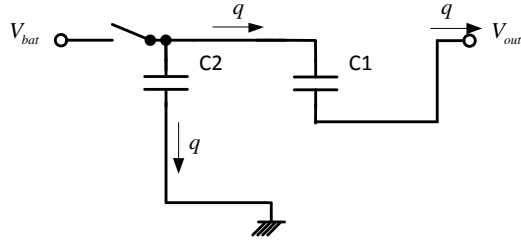
Charge transfer loss exists due to the fact that when two capacitors with different voltage potential are connected, there will be energy loss due to the charge redistribution. To see this, consider the 2:1 SC-VRM in Figure 1.9(c). We can calculate its charge transfer loss by calculating the difference between the input and output energy. We proceed using the charge multiplier method. Figure 2.2(a) and 2.2(b) show the charge multiplier in the first and second phase. The total input and output energy during one converter cycle



(a)



(b)



(c)

Figure 2.1: Fibonacci SC-VRM: (a) VRM architecture, (b) phase 1 configuration and (c) phase 2 configuration.

can be expressed as:

$$E_{bat} = qV_{bat} \quad (2.2)$$

$$E_{out} = 2qV_{dd} \quad (2.3)$$

Subtracting (2.3) from (2.2), we arrive at the energy loss expression:

$$E_{loss} = E_{bat} - E_{core} = q(V_{bat} - 2V_{dd}) \quad (2.4)$$

Since  $q$  is the charge transferred from the input,  $q$  can be expressed as a function of the voltage variation on the flying capacitor  $C_{sc}$ :

$$q = C_{sc}\Delta V = C_{sc}(V_{bat} - 2V_{dd}) \quad (2.5)$$

Substituting (2.5) into (2.4) yields the final expression for the energy loss:

$$E_{loss} = C_{sc}(V_{bat} - 2V_{dd})^2 \quad (2.6)$$

Usually, the energy loss is modeled as an effective series resistor (ESR). To derive an analytical expression for the ESR, we need to relate the energy loss to the load current. Assuming that  $f_{sw}$  is the switching frequency, the load current  $I_{out}$  is given by the charge transferred per cycle divided by the converter switching period:

$$I_{out} = 2C_{sc}(V_{bat} - 2V_{dd})f_{sw} \quad (2.7)$$

Therefore, the ESR can be expressed as:

$$R = \frac{E_{loss}}{I_{out}^2/f_{sw}} = \frac{1}{4C_{sc}f_{sw}} \quad (2.8)$$

In the above derivation, we make an assumption that the converter is operating in slow switching limit (SSL) [11], which implies that the switching frequency is much slower than the RC time constant of the switch on-resistance and the charge transfer capacitor. This is a reasonable assumption because the power level this thesis considers is relatively low ( $\mu\text{W}$  -  $\text{mW}$  range) so conduction loss is small.

To get the SSL resistance for an arbitrary SC-VRM, we use the charge multiplier method and Tellegen's theory [12]. Assume that in a two-phase SC-VRM, there are  $k$  charge transfer capacitors,  $C_i (i = 1, \dots, k)$ , and  $k$  charge multipliers  $\beta_i (i = 1, \dots, k)$ . Let  $\beta_{in}$ ,  $\beta_{out}$  denote the charge multipliers of input and output voltage source, respectively. Employing Tellegen's theory, the inner product of charge multiplier vector and the corresponding capacitor voltage vector should be zero.

In phase one:

$$\sum_{i=1}^k V_{i,1}\beta_{i,1} + V_{in}\beta_{in,1} + V_{dd}\beta_{out,1} = 0 \quad (2.9)$$



In phase two:

$$\sum_{i=1}^k V_{i,1}\beta_{i,2} + V_{in}\beta_{in,2} + V_{dd}\beta_{out,2} = 0 \quad (2.10)$$

For a two-phase SC-VRM, under steady state, the charge flowing into a capacitor in one phase must be equal to the charge supplied by it in the other phase. Therefore we have  $\beta_{i,1} = -\beta_{i,2}$ . Subtracting (2.9) from (2.10) we get:

$$\sum_{i=1}^k (V_{i,1} - V_{i,2}) |\beta_i| + V_{dd}\beta_{out} = V_{in}\beta_{in} \quad (2.11)$$

The energy loss expression, for an arbitrary two phase SC-VRM, is therefore given by:

$$E_{loss} = E_{bat} - E_{core} = \sum_{i=1}^k (V_{i,1} - V_{i,2}) |\beta_i| = \sum_{i=1}^k \frac{|\beta_i|^2}{C_i} \quad (2.12)$$

On the other hand, the energy loss can be derived from the input and output charge multiplier:

$$E_{loss} = \left(\frac{V_{in}}{N} - V_{dd}\right)\beta_{out} \quad (2.13)$$

where  $N$  denotes the ideal conversion ratio. Equating the above two equations we arrive at an expression for the output voltage:

$$V_{dd} = \frac{V_{in}}{N} - \sum_{i=1}^k \frac{|\beta_i|^2}{C_i\beta_{out}} = \frac{V_{in}}{N} - \sum_{i=1}^k \frac{|\beta_i|^2}{C_i f_{sw} \beta_{out}^2} I_{out} \quad (2.14)$$

From (2.14) we can see that SC-VRM can be modeled as a voltage source representing the ideal conversion output, in series with a SSL resistor which captures the charge transfer loss:

$$R_{SSL} = \frac{E_{loss}}{I_{load}^2 / f_{sw}} = \sum_{i=1}^k \frac{|\beta_i|^2}{\beta_{out}^2 C_i f_{sw}} \quad (2.15)$$

A more general expression for multiphase SSL resistance can be found in [11].

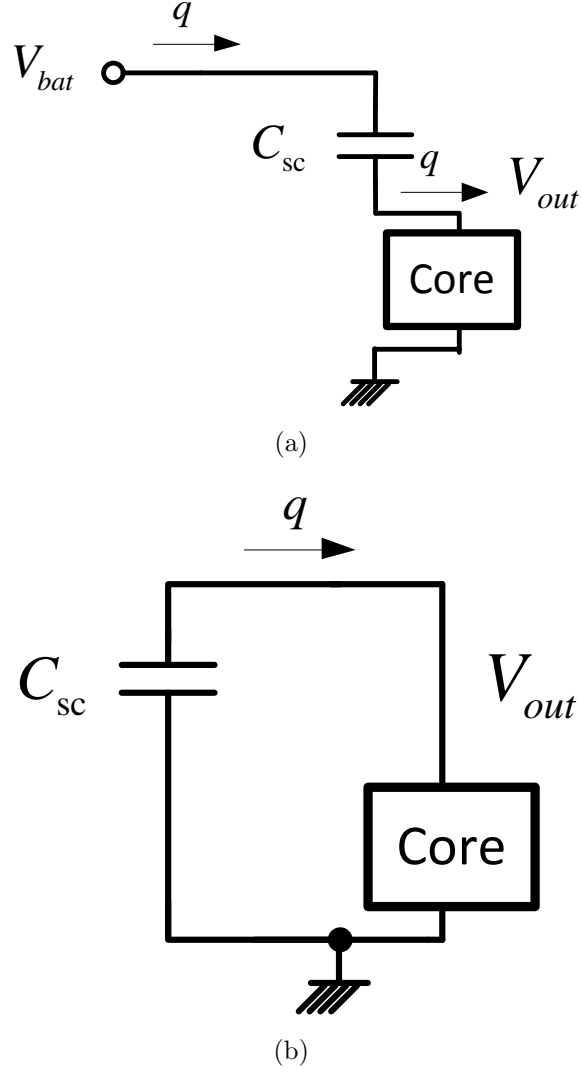


Figure 2.2: 2:1 SC-VRM charge transfer loss calculation using charge multiplier method: (a) connection in the first phase and (b) connection in the second phase.

### 2.2.2 Drive loss

To alter the connection for the capacitor array, a driver circuit is needed to control power switches which are usually made wide to reduce their on-resistance. Driving the large power switches will lead to an additional energy loss. Assuming that the total load capacitance from the power switches is  $C_{gate}$ , the drive loss can be expressed as:

$$E_{drive} = C_{gate} V_{bat}^2 \quad (2.16)$$

This suggests that the drive loss can be modeled as a shunt resistance between the input and ground, as follows:

$$R_{drive} = \frac{V_{bat}^2}{E_{drive} f_{sw}} = \frac{1}{C_{gate} f_{sw}} \quad (2.17)$$

Equation (2.17) shows that the drive loss increases with the size of the power switches ( $C_{gate}$ ) and switching frequency ( $f_{sw}$ ). To reduce the driver overhead, researchers have proposed frequency modulation [11] and dynamic transistor sizing [13] methods to scale the drive loss along with load current for higher efficiency.

### 2.2.3 Bottom plate capacitor loss

The bottom plate capacitor is the parasitic capacitor between the bottom plate of the on-chip capacitor and the substrate. It scales with the main capacitor area and can be as high as 5% of the main capacitor [14]. Since the bottom plate of the flying capacitor is not tied to a fixed voltage, the charge stored on the bottom plate capacitor may be wasted instead of being delivered to the output. For example, Figure 1.9(c) shows that the bottom plate capacitor of  $C_{sc}$  is charged and discharged every converter cycle. The bottom plate capacitor loss is given by:

$$E_{bp} = \gamma \sum_{i=1}^k C_i V_{bat}^2 = \gamma C_{tot} V_{bat}^2 \quad (2.18)$$

where  $C_{tot} = \sum_{i=1}^k C_i$  is the total charge transfer capacitances,  $C_i$ s are the individual charge transfer capacitances.  $\gamma$  denotes the ratio of bottom plate capacitances to  $C_{tot}$ . The bottom plate capacitor loss can also be modeled as a shunt resistance between input and ground, as follows:

$$R_{bp} = \frac{V_{bat}^2}{E_{bp} f_{sw}} = \frac{1}{\gamma \sum_{i=1}^k C_i f_{sw}} \quad (2.19)$$

Table 2.2: SC-VRM Energy Model

	Expression	Model
Charge transfer loss	$\sum_{i=1}^k \frac{ \beta_i ^2}{C_i}$	series resistor
Drive loss	$C_{gate} V_{bat}^2$	shunt resistor
Bottom plate capacitor loss	$\gamma \sum_{i=1}^k C_i V_{bat}^2$	shunt resistor
Control loss	$C_{ctrl} V_{bat}^2$	shunt resistor

#### 2.2.4 Control loss

The SC-VRM needs a control circuit to regulate the output voltage  $V_{dd}$ . As suggested in (2.14), we can achieve regulation for the SC-VRM by changing conversion ratio  $N$ , switching frequency  $f_{sw}$ , or the flying capacitor values  $C_i$ . In practice, frequency regulation is the most widely used method due to its simplicity and its ability to reduce driver overhead by using frequency modulation or pulse skipping. The control circuit contributes to energy loss that is independent of the load current. Therefore, it needs to be taken into consideration in the optimization procedure, especially under light load conditions. The control loss can be expressed as:

$$E_{ctrl} = C_{ctrl} V_{bat}^2 \quad (2.20)$$

where  $C_{ctrl}$  is the equivalent control circuit load capacitance. The control loss can be modeled by a shunt resistance similar to the drive loss and bottom plate capacitor loss, as follows:

$$R_{ctrl} = \frac{V_{bat}^2}{E_{ctrl} f_{ctrl}} = \frac{1}{C_{ctrl} f_{ctrl}} \quad (2.21)$$

where  $f_{ctrl}$  is the control circuit frequency. and is separated with  $f_{sw}$  to account for the load independence of control loss with frequency modulation control scheme. Table 2.2 summarizes the energy model of SC-VRM.

### 2.2.5 Conventional optimization

Conventional the design of SC-VRM focuses on optimizing its efficiency  $\eta$ , which can be formulated as follows:

$$\begin{array}{ll}
 \text{max.} & \eta(f_{sw}, C_{tot}, N) \\
 & = \frac{P_{core}}{P_{core} + P_{ctl} + P_{drive} + P_{bp} + P_{ctrl}} \\
 \text{s.t.} & C_{tot} \leq C_{max}
 \end{array} \tag{2.22}$$

The above optimization problem treats the core as a load resistor  $R_{core}$  drawing a specific load current  $I_{core}$  and needing a voltage of  $V_{dd} = I_{core}R_{core}$ . VRM is designed to optimize the ratio of VRM's output power  $E_{core}$  and total power drawn from the battery. The efficiency optimum design point, defined as the tuple  $(V_{dd,\eta}^*, I_{core}^*)$  is obtained by balancing the charge transfer loss  $E_{ctl}$  (series loss) with the shunt losses (sum of  $E_{bp}$ ,  $E_{drive}$ ,  $E_{ctrl}$ ).

## CHAPTER 3

# JOINT OPTIMIZATION OF THE SC-VRM AND THE CORE

In the previous chapter, an energy model is developed for a general SC-VRM topology, and is summarized in Figure 3.1. An ideal SC-VRM can be modeled as a DC transformer, with conversion ratio  $N:1$ . Charge transfer loss is modeled as a series resistor  $R_{SSL}$  between the transformer output and the actual output node. Other losses, such as drive loss ( $R_{drive}$ ), bottom plate capacitor loss ( $R_{bp}$ ) and control loss ( $R_{ctrl}$ ) can all be modeled as shunt resistors between input and ground. The simple model in Figure 3.1 captures the fundamental loss components in any SC-VRM and can be useful in the following optimization procedure. The design parameters in the model are the capacitor values ( $C_i$ ), switch sizes ( $C_{gate}$ ), the switching frequency ( $f_{sw}$ ), and the output voltage ( $V_{dd}$ ). Conventional SC-VRM optimization employs efficiency  $\eta$  as the cost function and optimizes for  $\eta$ , treating the core as a resistor or current source. The drawback of the  $\eta$ -oriented optimization method is that it treats the design of the core and SC-VRM separately. In the core design, SC-VRM is assumed to generate a constant voltage and the core is optimized for minimum energy operation. In the design of SC-VRM, optimization is performed without the notion of core speed/throughput requirements. To solve this problem, we propose to jointly optimize SC-VRM and the core with throughput as additional constraints. We first formulate the joint optimization problem where input energy per instruction  $E_{bat}$  is used as the cost function. Then we solve the problem by first determining how to size capacitors and switches, followed by system level optimization to minimize  $E_{bat}$ . We also present the comparison between the results of the proposed joint optimization and  $\eta$ -oriented optimization. Finally, we extend the joint optimization to reconfigurable SC-VRM and study the optimum conversion ratio setting strategy.

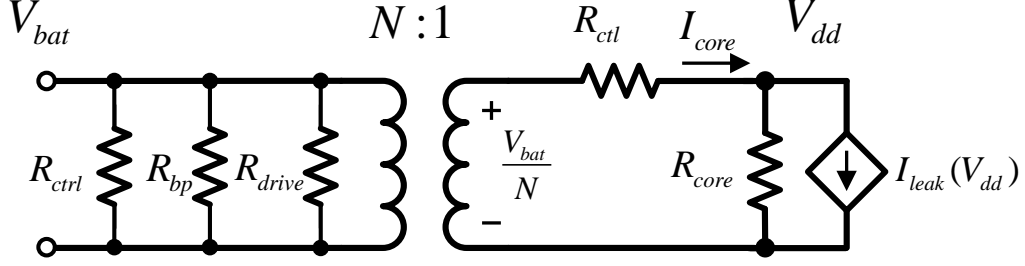


Figure 3.1: SC-VRM energy model.

### 3.1 Core-aware SC-VRM Energy Model

We propose a core-aware SC-VRM energy model, which differs from the conventional SC-VRM energy model in two respects: 1) core throughput/clock frequency  $f_{clk}$  is introduced as an additional parameter, and 2) the energy loss is normalized with respect to the core clock period  $T_{clk} = 1/f_{clk}$ . Thus, 1) identifies a range of feasible output voltages  $V_{dd}$  for the SC-VRM to generate, and 2) enables the use of energy per instruction (EPI) as a metric for optimization. For example, the clock frequency of the core determines the minimum required supply voltage  $V_{dd,min}(f_{clk})$ . Hence, the feasible set of SC-VRM output voltages  $V_{dd} \geq V_{dd,min}(f_{clk})$ . Thus, 1) identifies a range of feasible output voltages  $V_{dd}$  for the SC-VRM to generate, and 2) enables the use of energy per instruction (EPI) as a metric for optimization. For example, the clock frequency of the core determines the minimum required supply voltage  $V_{dd,min}(f_{clk})$ . Hence, the feasible set of SC-VRM output voltages  $V_{dd} \geq V_{dd,min}(f_{clk})$ .

To account for system energy, core EPI should also be considered. We adopt the core model presented in Chapter 1, which is repeated as follows:

$$E_{core} = C_{core} V_{dd}^2 + V_{dd} I_{leak}(V_{dd}) \frac{1}{f_{clk}} \quad (3.1a)$$

with

$$I_{leak}(V_{dd}) = \mu C_{ox} \frac{W}{L} (m-1) V_T^2 e^{\frac{-V_t}{mV_T}} e^{\frac{-\eta_d V_{dd}}{mV_T}} (1 - e^{\frac{-V_{dd}}{V_T}}) \quad (3.1b)$$

where  $C_{core}$  is the effective load capacitance in the core,  $V_{dd}$  is the supply voltage,  $V_t$  is threshold voltage,  $V_T$  is thermal voltage,  $\mu$  is the carrier mobility,  $C_{ox}$  is the gate capacitance per unit  $W/L$ ,  $m$  is a constant related

with sub-threshold slope factor, and  $\eta_d$  is the drain induced barrier lowering (DIBL) coefficient. This model captures the trade-off between the dynamic and leakage energy which leads to the well-studied *MEOP* [4]. The core is modeled as a load resistor  $R_{core}$  in parallel with a leakage current source  $I_{leak}(V_{dd})$  (see Figure 3.1).

Applying KCL at the output node of the SC-VRM model in Figure 3.1 indicates that its output voltage  $V_{dd}$  satisfies the following constraint:

$$I_{core} = \beta C_{tot} \left( \frac{V_{bat}}{N} - V_{dd} \right) f_{sw} = C_{core} V_{dd} f_{clk} + I_{leak}(V_{dd}) \quad (3.2)$$

We can see that output voltage  $V_{dd}$  and its dependence on the core's throughput  $f_{clk}$ , links the SC-VRM and core design parameter spaces.

We further assume that the power switches are sized such that the  $RC$  time constant is much ( $M > 1$  times) smaller than the SC-VRM switching period  $1/f_{sw}$ , i.e.,

$$M = \frac{R_{on} C_{tot}}{M f_{sw}} \quad (3.3)$$

where  $M$  is typically chosen to be between 3 and 5. The effective output capacitance of the drive circuit can then be expressed as:

$$C_{drive} = C_{gate} R_{on} M f_{sw} C_{tot} = \frac{M f_{sw}}{f_t} C_{tot} \quad (3.4)$$

where  $C_{gate}$  is the gate capacitance of the switch, and  $f_t$  is the unity gain frequency of the process technology.

Thus, the energy loss components per instruction for both the SC-VRM and the core can be obtained as follows:

$$E_{ctl} = \beta C_{tot} \left( \frac{V_{bat}}{N} - V_{dd} \right)^2 \frac{f_{sw}}{f_{clk}} \quad (3.5a)$$

$$E_{drive} = C_{tot} V_{bat}^2 \frac{M f_{sw}^2}{f_t f_{clk}} \quad (3.5b)$$

$$E_{bp} = \gamma C_{tot} V_{bat}^2 \frac{f_{sw}}{f_{clk}} \quad (3.5c)$$

$$E_{ctrl} = C_{ctrl} V_{bat}^2 \frac{f_{ctrl}}{f_{clk}} \quad (3.5d)$$



## 3.2 System-level Energy Optimization

We formulate the joint system-level optimization problem below:

$$\begin{array}{ll}
\min. & E_{bat}(f_{sw}, C_{tot}, N) \\
& = E_{core} + E_{ctl} + E_{drive} + E_{bp} + E_{ctrl} \\
\text{s.t.} & C_{tot} \leq C_{max} \\
& V_{dd} \geq V_{dd, \min}(f_{clk}) \\
& \beta C_{tot} \left( \frac{V_{bat}}{N} - V_{dd} \right) f_{sw} = C_{core} V_{dd} f_{clk} + I_{leak}(V_{dd})
\end{array} \tag{3.6}$$

where we assume that the core architecture and the battery voltage  $V_{bat}$  are fixed. Note: the key differences between (3.6) and (2.22) are in the objective function and the additional constraints imposed by the core throughput. We solve this optimization problem by first proving the following two results: 1) the optimum value of  $C_{tot}$  ( $C_{tot}^*$ ) is equal to  $C_{max}$ , and 2) the optimum  $f_{sw}$  ( $f_{sw}^*$ ) is the smallest value for which  $V_{dd} = V_{dd, \min}(f_{clk})$ . These results are a direct outcome of imposing a constraint on the core throughput  $f_{clk}$ .

This is a multi-variable optimization problem. We first solve the fixed conversion ratio (i.e. fixed  $N$ ) optimization problem by solving two partial minimization problems: 1) for fixed  $f_{sw}$ , determine the optimal  $C_i$ s, and 2) based on the optimal  $C_i$ s, determine the optimum switching frequency  $f_{sw}$ . After solving the fixed ratio optimization, we extend the joint optimization to reconfigurable SC-VRM.

### 3.2.1 Capacitor optimization

First, for fixed conversion ratio  $N$ , switching frequency  $f_{sw}$  and charge transfer capacitance  $C_{tot}$ , we solve for the optimal  $C_i$  to minimize  $E_{bat}$ . We begin by first simplifying the  $E_{bat}$  expression in (3.6). Substitute (3.2) and (3.4) into (3.6), we can simplify  $E_{bat}$  as:

$$\begin{aligned}
E_{bat} = & \frac{1}{\frac{1}{C_{core} f_{clk}} + \sum_{i=1}^k \frac{|\beta_i|^2}{\beta_{out}^2 C_i f_{sw}}} \left( \frac{V_{bat}}{N} \right)^2 \frac{1}{f_{clk}} \\
& + \left( M \frac{f_{sw}^2}{f_t f_{clk}} + \gamma \frac{f_{sw}}{f_{clk}} \right) V_{bat}^2 C_{tot} + V_{bat}^2 C_{ctrl} \frac{f_{ctrl}}{f_{clk}}
\end{aligned} \tag{3.7}$$

To solve for optimal  $C_i$ , we employ Lagrange multiplier method since the objective function has a positive semidefinite Hessian and therefore is convex.

$$L(\lambda) = \frac{1}{\frac{1}{C_{core}f_{clk}} + \sum_{i=1}^k \frac{|\beta_i|^2}{\beta_{out}^2 C_i f_{sw}}} \left( \frac{V_{bat}}{N} \right)^2 \frac{1}{f_{clk}} + \lambda \left( \sum_{i=1}^k C_i - C_{tot} \right) \quad (3.8)$$

$$+ (M \frac{f_{sw}^2}{f_t f_{clk}} + \gamma \frac{f_{sw}}{f_{clk}}) V_{bat}^2 C_{tot} + V_{bat}^2 C_{ctrl} \frac{f_{ctrl}}{f_{clk}} \quad (3.9)$$

The optimal capacitor value can be calculated by setting the partial derivative of the Lagrange with respect to each charge transfer capacitor to 0:

$$\frac{\partial L(\lambda)}{\partial C_i} = \frac{1}{\left( \frac{1}{C_{core}f_{clk}} + \sum_{i=1}^k \frac{|\beta_i|^2}{\beta_{out}^2 C_i f_{sw}} \right)^2} \frac{1}{\beta_{out}^2 f_{sw}} \frac{|\beta_i|^2}{C_i^2} \left( \frac{V_{bat}}{N} \right)^2 \frac{1}{f_{clk}} + \lambda = 0 \quad (3.10)$$

Solving the above equation, we get the optimal capacitor value and the corresponding energy loss:

$$C_i^* = \frac{|\beta_i|}{\sum_{i=1}^k |\beta_i|} C_{tot} \quad (3.11)$$

$$E_{bat}^* = \frac{1}{\frac{1}{C_{core}f_{clk}} + \frac{(\sum_{i=1}^k |\beta_i|)^2}{\beta_{out}^2 C_{tot} f_{sw}}} \left( \frac{V_{bat}}{N} \right)^2 \frac{1}{f_{clk}} \quad (3.12)$$

$$+ (M \frac{f_{sw}^2}{f_t f_{clk}} + \gamma \frac{f_{sw}}{f_{clk}}) V_{bat}^2 C_{tot} + V_{bat}^2 C_{ctrl} \frac{f_{ctrl}}{f_{clk}} \quad (3.13)$$

It is shown in ( 3.11) that the optimal capacitor value is proportional to the square of the charge multiplier that flows in/out of it. Intuitively, this means that we should assign a larger capacitor to the branch where current density is higher.

We now solve for the optimum total charge transfer capacitance  $C_{tot}$ .

**Lemma 1** *The product  $f_{sw}C_{tot}$  is a constant for a fixed conversion ratio  $N$ , core throughput  $f_{clk}$ , and the SC-VRM output voltage  $V_{dd}$ .*

## PROOF

From (3.2), we obtain:

$$C_{tot}f_{sw} = \frac{C_{core}V_{dd}f_{clk} + I_{leak}(V_{dd})}{\beta(\frac{V_{bat}}{N} - V_{dd})} \quad (3.14)$$

Thus, the product  $f_{sw}C_{tot}$  is a constant as all the variables on the R.H.S. of (3.14) are constants. This completes the proof.

Lemma 1 indicates that the input energy  $E_{bat}$  decreases as  $C_{tot}$  increases at a given output voltage  $V_{dd}$ .

**Theorem 1** *The input energy  $E_{bat}$  is a monotonically decreasing function of  $C_{tot}$  for fixed conversion ratio  $N$ , core throughput  $f_{clk}$ , and the SC-VRM output voltage  $V_{dd}$ .*

## PROOF

We define series energy dissipated  $E_{ser}$  in the combined SC-VRM and core (Figure 3.1) as:

$$E_{ser} = E_{ctl} + E_{core} = I_{core} \frac{V_{bat}}{N} \frac{1}{f_{clk}} \quad (3.15)$$

which scales with the load current  $I_{core}$ . Substituting (3.2) in (3.15), we obtain:

$$E_{ser} = \frac{V_{bat}}{N} \frac{C_{core}V_{dd}f_{clk} + I_{leak}(V_{dd})}{f_{clk}} \quad (3.16)$$

Equation (3.16) indicates that  $E_{ser}$  is constant for a fixed values of  $N$ ,  $f_{clk}$ , and  $V_{dd}$ . This implies that  $E_{ctl}$ , a SC-VRM parameter, is completely determined by the core throughput and complexity ( $C_{core}$ ). Similarly, the SC-VRM shunt loss  $E_{shunt}$  is given by:

$$\begin{aligned} E_{shunt} &= E_{drive} + E_{bp} + E_{ctrl} \\ &= \left(\frac{M}{f_t}(C_{tot}f_{sw}^2) + \gamma C_{tot}f_{sw} + C_{ctrl}f_{ctrl}\right) \frac{V_{bat}^2}{f_{clk}} \end{aligned} \quad (3.17)$$

Equation (3.17) indicates that for any  $f_{sw1} > f_{sw2}$ ,  $E_{shunt1} > E_{shunt2}$ . Therefore,  $E_{shunt}$  is an increasing function of  $f_{sw}$ . Both Recall from Lemma 1 that the product  $C_{tot}f_{sw}$  is a constant. Thus,  $E_{shunt}$  is a decreasing function of  $C_{tot}$ . Since  $E_{ser}$  is a constant, and  $E_{bat} = E_{ser} + E_{shunt}$ , it is clear that  $E_{bat}$  is a decreasing function of  $C_{tot}$ .

The following corollary is stated without proof:

**Corollary 1**

$$C_{tot}^* = C_{max} \quad (3.18)$$

### 3.2.2 Switching frequency optimization

Next, we prove that the input energy  $E_{bat}$  is an increasing function of  $f_{sw}$  starting with the following Lemma:

**Lemma 2** *The input energy  $E_{bat}$  and the SC-VRM output voltage  $V_{dd}$  are increasing functions of  $f_{sw}$  for a fixed conversion ratio  $N$ , and core throughput  $f_{clk}$ .*

**PROOF**

From (3.2),  $f_{sw}$  can be expressed as:

$$f_{sw} = \frac{C_{core}V_{dd}f_{clk} + I_{leak}(V_{dd})}{\beta(\frac{V_{bat}}{N} - V_{dd})C_{tot}} \quad (3.19)$$

Substituting  $C_{tot} = C_{max}$  from Corollary 1, we obtain:

$$f_{sw} = \frac{C_{core}V_{dd}f_{clk} + I_{leak}(V_{dd})}{\beta(\frac{V_{bat}}{N} - V_{dd})C_{max}} \quad (3.20)$$

Equation (3.20) indicates that the numerator increases while its denominator decreases as  $V_{dd}$  increases. This proves that  $f_{sw}$  is a strictly increasing function of  $V_{dd}$ . Since the inverse of a strictly increasing function exists and is also an increasing function, we conclude that  $V_{dd}$  is an increasing function of  $f_{sw}$ . Next, we show that  $E_{bat}$  is also an increasing function of  $f_{sw}$ . From (3.16), we see that  $E_{ser}$  is an increasing function of  $V_{dd}$ , and thus an increasing function of  $f_{sw}$  from to Lemma 2. Next, we substitute (3.18) into (3.17) to obtain:

$$E_{shunt} = (\frac{M}{f_t}(C_{max}f_{sw}^2) + \gamma C_{max}f_{sw} + C_{ctrl}f_{ctrl})\frac{V_{bat}^2}{f_{clk}} \quad (3.21)$$

which shows that  $E_{shunt}$  increases as  $f_{sw}$  becomes higher. Therefore, the input energy  $E_{bat} = E_{ser} + E_{shunt}$  is an increasing function of  $f_{sw}$ .

Lemma 2 can be used to prove the following theorem:

**Theorem 2** *The energy optimum SC-VRM output voltage  $V_{dd}^*$  is the lowest supply voltage needed by the core to meet the throughput requirement of  $f_{clk}$ , and the energy-optimum SC-VRM switching frequency  $f_{sw}^*$  is lowest switching frequency that enables the SC-VRM to generate  $V_{dd}^*$ , for a fixed conversion ratio  $N$  and core throughput  $f_{clk}$ .*

**PROOF**

The proof follows from Lemma 2.

### 3.3 Simulation Results

In this section, we first extract the core-aware SC-VRM energy model and then employ this model to jointly optimize the SC-VRM and the core.

#### 3.3.1 Core-aware SC-VRM energy model extraction

To verify the model, a 2:1 ladder SC-VRM was designed to the circuit schematic level in a commercial 1.2 V 130 nm process with a  $C_{tot} = 500$  pF. A pulse frequency modulated driver circuit consisting of a current starved oscillator, a non-overlap clock generator with embedded buffer chain, and a double comparator based hysteresis regulation loop [11] is designed to regulate the SC-VRM output voltage  $V_{dd}$  over a range of 400 mV to 580 mV. The purpose of the pulse frequency control method is to reduce the drive loss and bottom plate capacitor loss under light load conditions. The core is a simple 50 stage ring oscillator. Circuit level simulation with HSPICE is performed over a wide load range to verify the core-aware energy model in Section 3.1. From circuit simulation results, we estimate  $\gamma$  to be 1%, and  $C_{ctrl}$  to be 2% of the main load capacitance.

Figure 3.2 compares the results of circuit simulation and the model (3.5d) for a typical subthreshold load current  $I_{core}$  range of  $50 \mu\text{A}$  to  $1 \text{ mA}$  [15]. From Figure 3.2, we can see that the model and simulation results match to within 5% under moderate ( $400 \mu\text{A}$ ) to maximum load ( $1 \text{ mA}$ ) load conditions. Under light load conditions, the maximum error is within 15% due to the unmodeled transient and leakage energy in the SC-VRM.

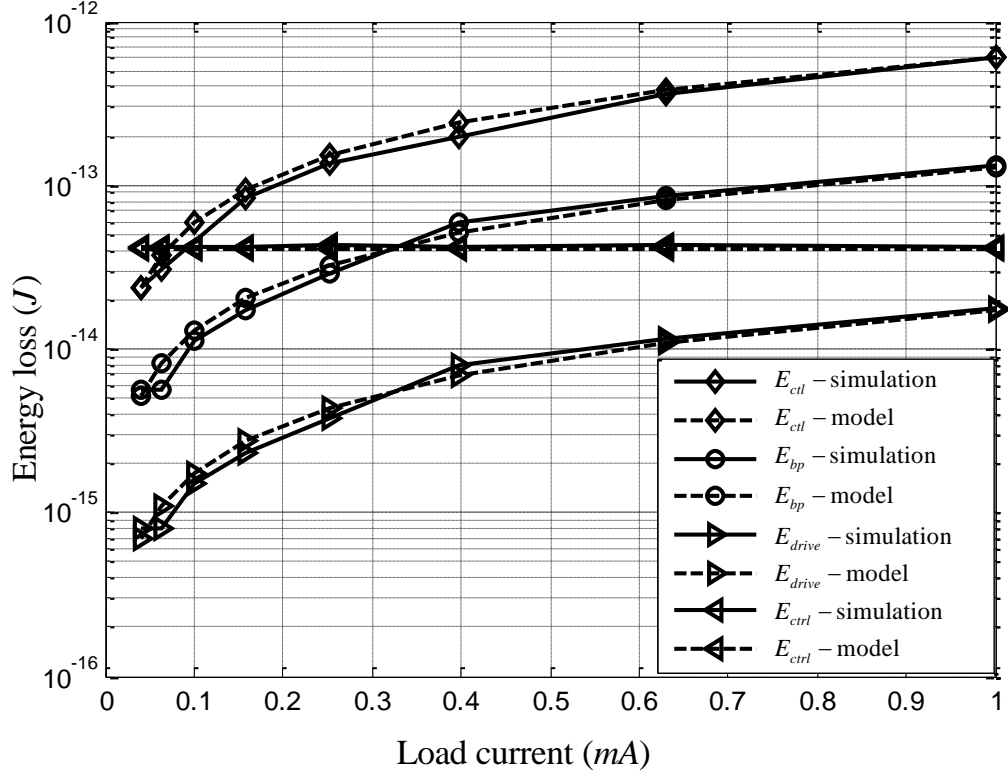


Figure 3.2: Comparison between HSPICE simulation results and analytical modeling results of a Ladder 2:1 SC-VRM shows that proposed model is valid over a wide load condition.

### 3.3.2 System-level optimization

To show the benefits of the proposed optimization, we compare the results of the conventional (2.22) and the proposed joint optimization (3.6) methods in Figure 3.3 for  $f_{clk} = 10$  MHz and  $C_{core} = 200$  pF. We find that the minimum  $E_{bat}$  for joint optimization  $E_{bat,sys}^* = 47.3$  pJ occurs at  $f_{sw,sys}^* = 1.4$  MHz for a  $V_{dd,sys}^* = 350$  mV, while the conventional approach results in an  $E_{bat,\eta}^* = 77.5$  pJ occurring at  $f_{sw,\eta}^* = 10$  MHz for a  $V_{dd,\eta}^* = 546$  mV, leading to a 38.9% reduction in energy. The conventional approaches maximize the VRM efficiency by balancing the series and shunt losses,  $E_{ser}$  and  $E_{shunt}$ , which leads to a larger value of  $f_{sw}$  and hence  $V_{dd}$  than is needed by the core. We can also see that  $E_{bat}$  increases with  $f_{sw}$  as expected because VRM losses increase, and because  $E_{core}$  increases due to higher output voltage  $V_{dd}$ .

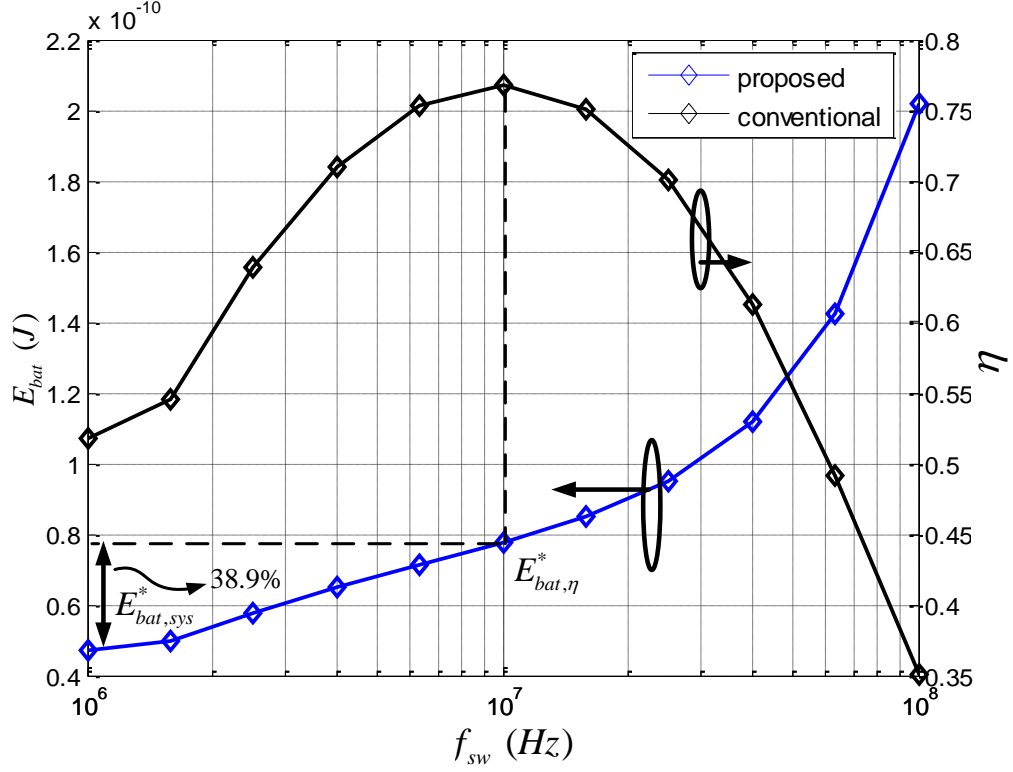


Figure 3.3: Comparison of the proposed and conventional optimization techniques for  $C_{core} = 200$  pF and  $f_{clk} = 10$  MHz.

Figure 3.4 plots the energy savings  $\Delta E_{bat}^*$  given by

$$\Delta E_{bat}^* = (E_{bat,sys}^* - E_{bat,\eta}^*) / E_{bat,\eta}^* \quad (3.22)$$

where  $E_{bat,sys}^*$  is the input energy at system optimum point, and  $E_{bat,\eta}^*$  is the input energy at efficiency maximum point. We see that the energy savings peaks in the medium throughput range of 20 MHz to 40 MHz where it can be as high as 38.9%. The energy savings reduce at lower throughputs because the efficiency-maximized supply voltage  $V_{dd,\eta}^*$  shifts towards the system optimum voltage  $V_{dd,sys}$ . At high throughputs, the system optimum point  $V_{dd,sys}$  moves to higher  $V_{dd}$  and as a result, the difference between  $V_{dd,sys}$  and  $V_{dd,\eta}$  reduces, leading to lower energy saving.

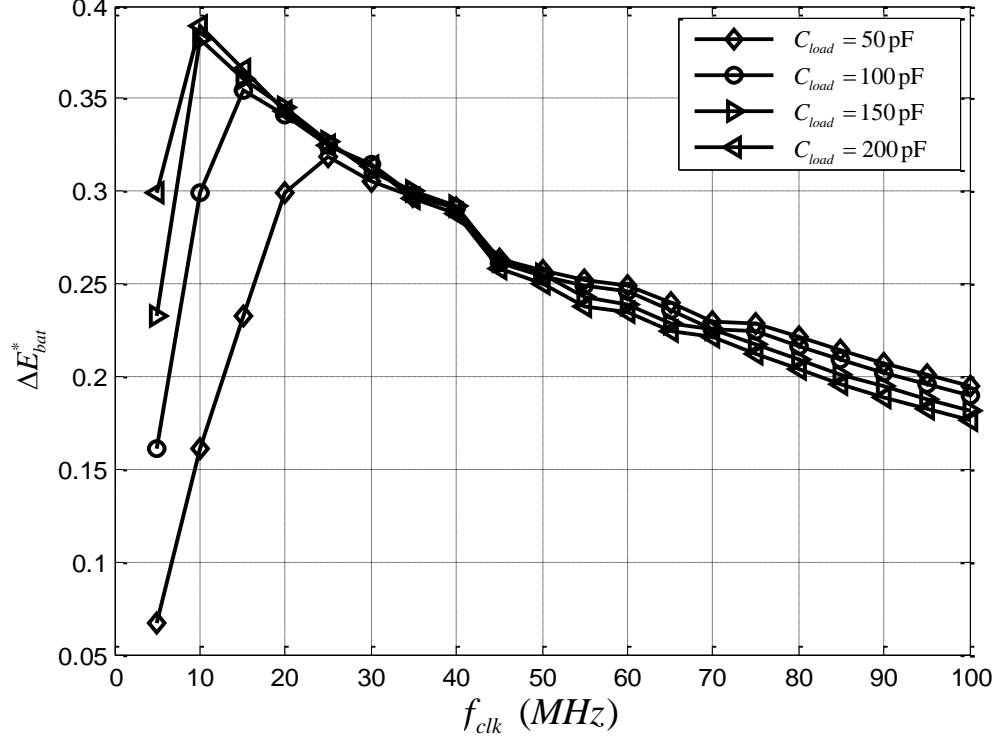


Figure 3.4: Energy savings as a function of  $f_{clk}$  for different  $C_{core}$ .

### 3.3.3 Reconfigurable SC-VRM Optimization

The SC-VRM has a drawback that its conversion ratio is fixed, and that it retains high efficiency for a narrow range of output voltages, which is contrary to the wide  $V_{dd}$  range requirement in DVFS systems. The reconfigurable SC-VRM solves this problem by configuring the capacitor switching network topology. In this section, we study the results of joint optimization for reconfigurable SC-VRMs. For simplicity, we employ a simple series-parallel reconfigurable configuration, as shown in Figure 3.5. The main capacitor is separated into three capacitors, such that a conversion ratio of 2:1, 3:1 and 4:1 can be easily obtained by selectively combining the three capacitors. To evaluate the effect of conversion ratio on input energy, we perform joint optimization for the three conversion ratios for various throughput requirements. For each throughput requirement  $f_{clk}$ , joint optimization is carried out to obtain the optimum  $f_{sw,sys}^*$ ,  $V_{dd,sys}^*$  and  $E_{bat,sys}^*$ .

Figure 3.6 shows  $E_{bat}$  as a function of throughput requirements for various conversion ratios. As shown in the figure, the energy-optimum conversion ratio depends upon the throughput requirements. For low throughputs, a



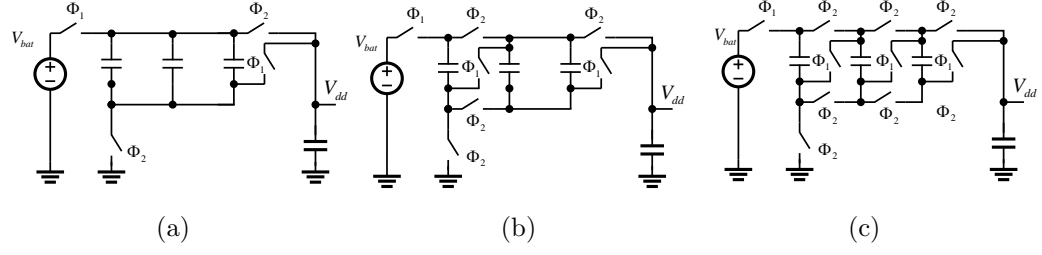


Figure 3.5: A reconfigurable SC-VRM.

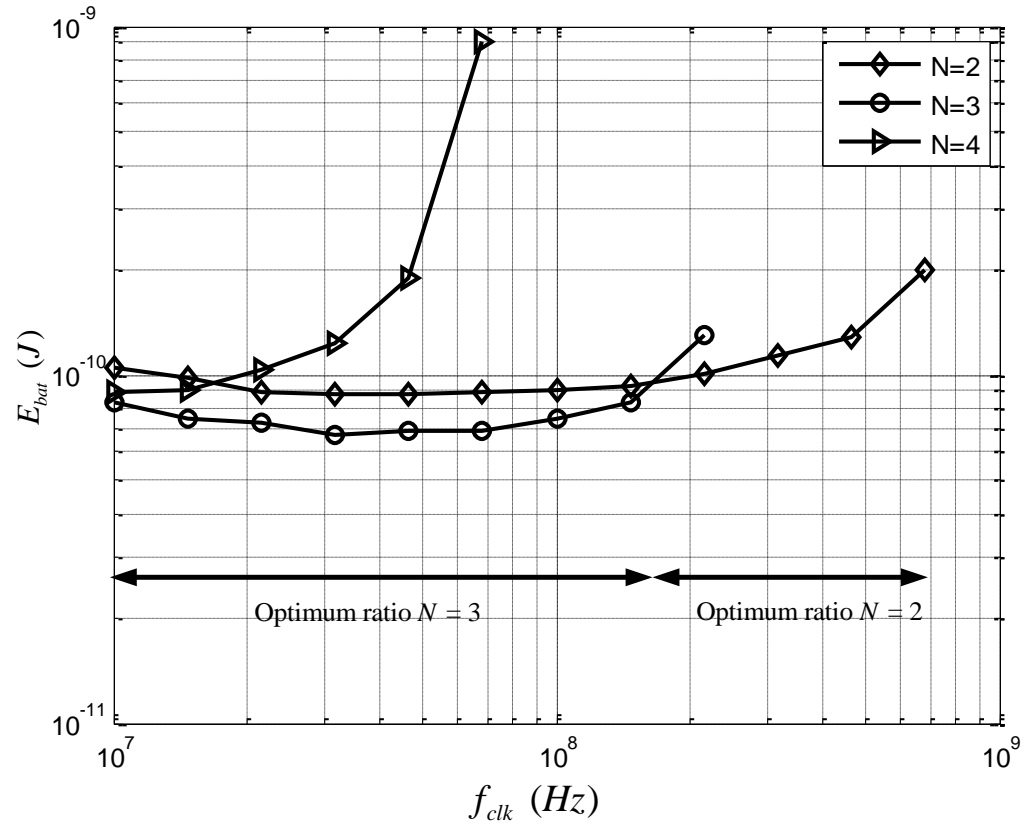


Figure 3.6: Joint optimization results for a reconfigurable SC-VRM with  $C_{core} = 200$  pF.

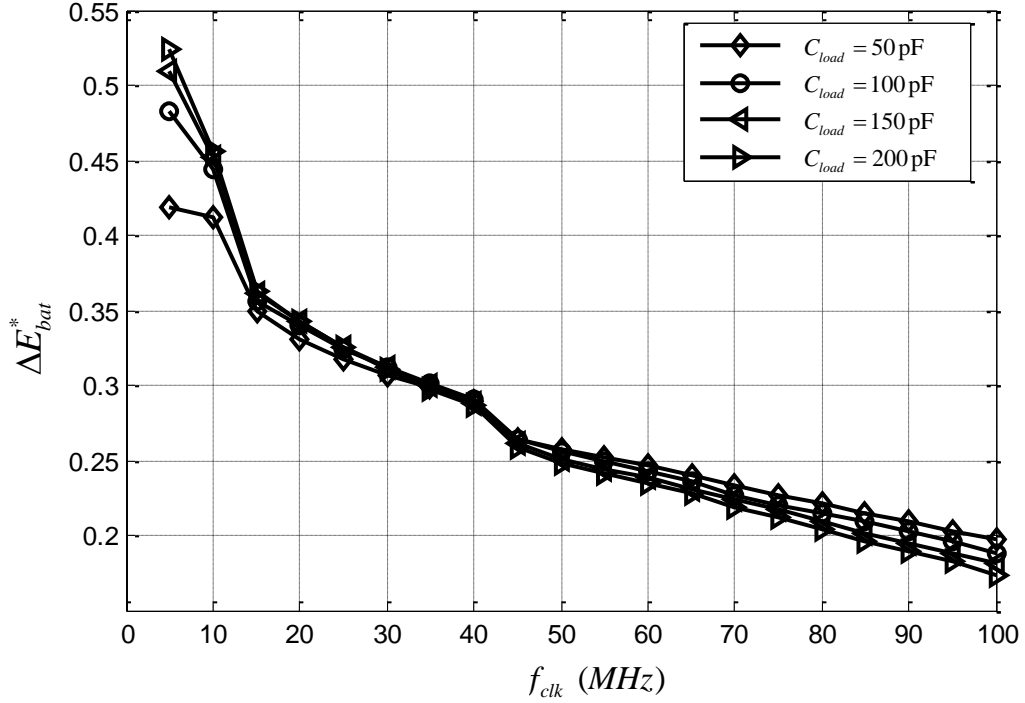


Figure 3.7: Energy saving of joint optimization with respect to efficiency oriented optimization for reconfigurable SC-VRM with conversion ratio of  $N=2,3$  and 4.

high conversion ratio is beneficial since it reduces the charge needed per computation. On the other hand, for high throughputs, a low conversion ratio is beneficial as it reduces the switching frequency  $f_{sw}$  significantly by transferring more charge per SC-VRM cycle. Therefore, optimum conversion ratio should be set according to throughput requirement to minimize system energy. Figure 3.7 shows the energy saving as defined in (3.22) with reconfigurable SC-VRM. For each throughput requirement, joint optimization is performed to select the optimum conversion ratio and compared with efficiency maximum design. From the figure, we can see in low throughput conditions the energy saving is increased dramatically compared to the fixed conversion ratio case, due to selecting the optimum conversion ratio, and a maximum energy saving of 52% can be achieved.

## CHAPTER 4

## CONCLUSION

Fully integrated SC-VRM is a promising solution for the increasingly challenging on-chip power management in embedded systems. Unlike the conventional SC-VRM design approach where efficiency is employed as the sole metric, in this thesis we propose to jointly optimize the SC-VRM and the compute core to minimize the total system energy per instruction. We demonstrate that by introducing throughput constraints from the compute core, the optimum charge transfer capacitance is the maximum available on-chip capacitance, and the optimum switching frequency is the minimum frequency that generates a throughput-satisfying output voltage. Furthermore, we show that joint optimization can be extended to reconfigurable SC-VRM to further optimize for system input energy. Simulation results show that 8% to 38.9% energy saving is achieved by choosing optimal switching frequency. For reconfigurable SC-VRM, 15% to 52% energy saving can be obtained by dynamically selecting conversion ratio based on throughput requirement. In the future, we will perform system optimization for TUsAs, and study architecture level optimization techniques such as parallelization and pipelining to further reduce system energy.

## REFERENCES

- [1] G. Moore, “No exponential is forever: but “forever” can be delayed! [semiconductor industry],” in *Solid-State Circuits Conference, 2003. Digest of Technical Papers. ISSCC. 2003 IEEE International*, vol. 1, 2003, pp. 20 – 23.
- [2] J. Rabaey, “The swarm at the edge of the cloud - a new perspective on wireless,” in *VLSI Circuits (VLSIC), 2011 Symposium on*, June 2011, pp. 6 –8.
- [3] R. Dreslinski, M. Wieckowski, D. Blaauw, D. Sylvester, and T. Mudge, “Near-threshold computing: Reclaiming Moore’s law through energy efficient integrated circuits,” *Proceedings of the IEEE*, vol. 98, no. 2, pp. 253 –266, Feb. 2010.
- [4] B. Zhai, D. Blaauw, D. Sylvester, and K. Flautner, “Theoretical and practical limits of dynamic voltage scaling,” in *Design Automation Conference, 2004. Proceedings. 41st*, July 2004, pp. 868 –873.
- [5] B. Calhoun and A. Chandrakasan, “Characterizing and modeling minimum energy operation for subthreshold circuits,” in *Low Power Electronics and Design, 2004. ISLPED ’04. Proceedings of the 2004 International Symposium on*, Aug. 2004, pp. 90 –95.
- [6] M. Stojcev, M. Kosanovic, and L. Golubovic, “Power management and energy harvesting techniques for wireless sensor nodes,” in *Telecommunication in Modern Satellite, Cable, and Broadcasting Services, 2009. TELSIKS ’09. 9th International Conference on*, Oct. 2009, pp. 65 –72.
- [7] Y. Lee, G. Chen, S. Hanson, D. Sylvester, and D. Blaauw, “Ultra-low power circuit techniques for a new class of sub-mm<sup>3</sup> sensor nodes,” in *Custom Integrated Circuits Conference (CICC), 2010 IEEE*, Sept. 2010, pp. 1 –8.
- [8] M. Seeman, V. Ng, H.-P. Le, M. John, E. Alon, and S. Sanders, “A comparative analysis of switched-capacitor and inductor-based dc-dc conversion technologies,” in *Control and Modeling for Power Electronics (COMPEL), 2010 IEEE 12th Workshop on*, June 2010, pp. 1 –7.

- [9] M. Seok, D. Jeon, C. Chakrabarti, D. Blaauw, and D. Sylvester, "Pipeline strategy for improving optimal energy efficiency in ultra-low voltage design," in *Design Automation Conference (DAC), 2011 48th ACM/EDAC/IEEE*, June 2011, pp. 990–995.
- [10] C. Zheng, L. Su, and D. Ma, "A systematic USFG design approach for integrated reconfigurable switched-capacitor power converters," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 58, no. 11, pp. 2790–2800, Nov. 2011.
- [11] M. D. Seeman, "A design methodology for switched-capacitor dc-dc converters," Ph.D. dissertation, EECS Department, University of California, Berkeley, May 2009. [Online]. Available: <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-78.html>
- [12] C. Desoer and E. Kuh, *Basic Circuit Theory*, ser. International student edition. McGraw-Hill, 1984. [Online]. Available: <http://books.google.com/books?id=pT7ZTE8RlkUC>
- [13] S. Musunuri and P. Chapman, "Optimization of CMOS transistors for low power dc-dc converters," in *Power Electronics Specialists Conference, 2005. PESC '05. IEEE 36th*, June 2005, pp. 2151–2157.
- [14] R. Aparicio and A. Hajimiri, "Capacity limits and matching properties of integrated capacitors," *Solid-State Circuits, IEEE Journal of*, vol. 37, no. 3, pp. 384–393, Mar 2002.
- [15] J. Kwong, Y. Ramadass, N. Verma, and A. Chandrakasan, "A 65 nm sub- microcontroller with integrated SRAM and switched capacitor dc-dc converter," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 1, pp. 115–126, Jan. 2009.